

Hybrid Features Based Face Recognition Method Using PCA and SVM

Karthik.G

2nd year mtech Department of TCE
Dayananda Sagar College of Engineering
Bangalore,India
karthik.knocks@gmail.com

Sateesh Kumar.H.C

Associate professor Department of TCE
Dayananda Sagar College of Engineering
Bangalore,India
hcsateesh@gmail.com

Abstract – Face recognition is a biometric tool for authentication and verification. It is having both research and practical relevance. Face recognition not only makes hackers virtually impossible to steal one's "password", but also increases the user-friendliness in human-computer interaction. A facial recognition based verification system can further be deemed a computer application for automatically identifying or verifying a person in a digital image. The two common approaches employed for face recognition are analytic (local features based) and holistic (global features based) approaches with acceptable success rates. In this paper, we present an intelligent hybrid features based face recognition method. It combines the local and global approaches to produce a complete, a robust and high success rate face recognition system. The global features are computed using principal component analysis while the local features are computed configuring the central moment and Eigen vectors and the standard deviation of the eyes, nose and mouth segments of the human face as the decision support entities of Support Vector Machine.

Keywords – face recognition; analytic approach; holistic approach; hybrid features; central moment; eigen vectors; standard deviation; support vector machine;

I. INTRODUCTION

Biometrics refers to a science of analyzing human body parts for security purposes. The word biometrics is derived from the Greek words *bios* (life) and *metrikos* (measure). Biometric identification is becoming more popular of late owing to the current security requirements of society in the field of information, business, military, e-commerce and etc. In general, biometric systems process raw data in order to extract a template which is easier to process and store, but carries most of the information needed.

Face recognition is a nonintrusive method, and facial images are the most common biometric characteristics used by humans to make a personal recognition. Human faces are complex objects with features that can vary over time. However, we humans have a natural ability to recognize faces and identify person at the spur of the second. Of course, our natural recognition ability extends beyond face recognition too. Nevertheless, in the interaction between humans and machines, also commonly

known as Human Robot Interface [3] or Human Computer Interface (HCI), the machines are to be trained to recognize and identify and differentiate the human faces. There is thus a need to simulate recognition artificially in our attempts to create intelligent autonomous machines. Recently face recognition is attracting much attention in the society of network multimedia information access.

Basically, any face recognition system can be depicted by the following block diagram.

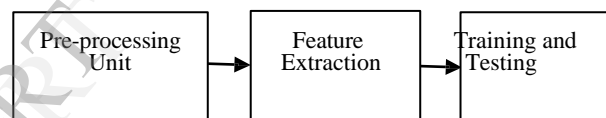


Figure 1. Basic blocks of a face recognition system.

1) *Pre-processing Unit*: In the initial phase, the image captured in the true colour format is converted to gray scale image and resized to a predefined standard and noise is removed. Further Histogram Equalization (HE) and Discrete Wavelet Transform (DWT) are carried out for illumination normalization and expression normalization respectively [4].

2) *Feature Extraction*: In this phase, facial features are extracted using Edge Detection Techniques, Principal Component Analysis (PCA) Technique, Discrete Cosine Transform (DCT) coefficients, DWT coefficients or fusion of different techniques [5].

3) *Training and Testing*: Here, Euclidean Distance (ED), Hamming Distance, Support Vector Machine (SVM), Neural Network [6] and Random Forest (RF) [7] may be used for training followed by testing the new images or the test images for recognition.

The popular approaches for face recognition are based either on the location and shape of facial attributes such as the eyes, eyebrows, nose, lips and chin, and their spatial relationships, or the overall analysis of the face image that represents a face as a weighted combination of a number of canonical faces.

In the former approach the local attributes of the face are considered in training and testing while the latter approach reckons in the information derived from the whole face. The local features based approach

demand a vast collection of database images for effective training thus increasing the computation time. The global technique works well with frontal view face images but they are sensitive to translation, rotation and pose changes [8].

Since, these two approaches do not give a complete representation of the facial image, hybrid features based face recognition system is designed which blends the two former approaches into a single system.

II. RELATED WORKS

The past few years have witnessed an increased interest in researches aiming at developing reliable face recognition techniques.

One of the commonly employed techniques involves representing the image by a vector in a dimensional space of size similar to the image [9]. However, the large dimensional space of the image reduces the speed and robustness of face recognition. This problem is overcome rather effectively by dimensionality reduction techniques such as the Principal Component Analysis (PCA) and the Linear Discriminant Analysis (LDA).

PCA is an eigenvector method designed to model linear variation in high-dimensional data. PCA performs dimensionality reduction by projecting an original n -dimensional data onto a k ($k \ll n$)-dimensional linear subspace spanned by the leading eigenvectors of the data's covariance matrix [10].

particle swarm optimization (PSO) is a computational paradigm based on the idea of collaborative behavior inspired by the social behavior of bird flocking or fish schooling. The algorithm is applied to coefficients extracted by two feature extraction techniques: the discrete cosine transforms (DCT) and the discrete wavelet transform (DWT). The proposed PSO-based feature selection algorithm is utilized to search the feature space for the optimal feature subset where features are carefully selected according to a well defined discrimination criterion.

While PCA uses orthogonal linear space for encoding information, LDA encodes using linearly separable space in which bases are not necessarily orthogonal. Experiments carried out by researchers thus far point to the superiority of algorithms based on LDA over PCA.

Another face analysis technique is the Locality Preserving Projections (LPP). It consists in obtaining a face subspace and finding the local structure of the manifold. Basically it is obtained by finding the optimal linear approximations to the Eigen functions of the Laplace Beltrami operator on the manifold. Therefore, it recovers important aspects of the intrinsic nonlinear manifold structure by preserving local structure though it is a linear technique [12].

Ramesha K and K B Raja, proposed Dual Transform based Feature Extraction for Face Recognition (DTBFEFR). Here Dual Tree Complex Wavelet Transform (DT-CWT) is employed to form the feature vector and Euclidean Distance (ED), Random Forest (RF) and Support Vector Machine (SVM) are used as the classifiers [13].

Weng and Huang presented a face recognition model based on hierarchical neural network which is grown automatically and not trained with gradient-descent. Good results for discrimination of ten distinctive subjects are reported [14].

This paper presents the face recognition method using both the geometrical features of the biometrical characteristic of the face such as eyes, nose, and mouth and the overall analysis of the whole face. After the pre-processing stage, segments of the eyes, nose and mouth are extracted from the faces of the database. These blocks are then resized and the training features are computed. These facial features reduce the dimensionality by gathering the essential information while removing all redundancies present in the segment. Besides, the global features of the total image are also computed. These specially designed features are then used as decision support entities of the classifier system configured using the SVM.

III. LOCAL AND GLOBAL FACE FEATURES EXTRACTION WITH MARKED DIMENSIONALITY REDUCTION

The purpose of feature extraction is to extract the feature vectors or information which represents the face and reduces computation time and memory storage.

Local facial feature extraction consists in localizing the most characteristic face components (eyes, nose, mouth, etc.) within images that depict human faces.

Global feature extraction consists in considering the face as a single whole entity and then extracting information provided by the whole face.

In this work, Central Moment, eigenvector of the eyes, nose and mouth are computed as the training features for the local feature extraction while standard deviation and eigenvector of the covariance of the whole face are assessed for the global features.

These features besides extracting the quintessential information of the face also account for dimensionality reduction.

A. Central Moment

Central moment finds its application in recognition of shape features which are independent of parameters and which cannot be controlled in an image are generated. Such features are called invariant features. There are several types of invariance. For example, if an object may occur in an arbitrary location in an image, then one needs the moments to be invariant to location. For binary connected components, this can be achieved simply by using the central moments, pq [15].

In image processing, computer vision and related fields, an image moment is a certain particular weighted average (moment) of the image pixels' intensities, or a function of such moments, usually chosen to have some attractive property or interpretation. Image moments are useful to describe objects after segmentation. Simple properties of the image which are found *via* image moments include area (or total intensity), its centroid, and information about its orientation [16].

Central moments are mathematically defined as [17]

$$\rho_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y) dx dy \quad (1)$$

$$\bar{x} = \frac{M_{10}}{M_{00}}, \quad \bar{y} = \frac{M_{01}}{M_{00}}$$

\bar{x} and \bar{y} are the components of the centroid. If $f(x, y)$ is a digital image, then the previous equation becomes

$$\rho_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y)$$

Central moments are translational invariant. Information about image orientation can be derived by first using the second order central moments to construct a covariance matrix.

The covariance matrix of the image $I(x, y)$ is

$$\text{cov}[I(x, y)] = \begin{bmatrix} m'_{20} & m'_{11} \\ m'_{11} & m'_{02} \end{bmatrix} \quad (4)$$

The eigenvectors of this matrix correspond to the major and minor axes of the image intensity, the orientation can thus be extracted from the angle of the eigenvector associated with the largest eigenvalue.

For higher order moments it is common to normalize these moments by dividing by m_0 (or m_{00}). This allows one to compute moments which depend only on the shape and not the magnitude of $f(x)$. The result of normalizing moments gives measures which contain information about the shape or distribution (not probability distribution) of $f(x)$. This is what makes moments useful for the analysis of shapes in image processing, for which $f(x, y)$ is the image function. These computed moments are usually used as features for shape recognition [18].

B. Eigenvector with Highest Eigen Value

An eigenvector of a matrix is a vector such that, if multiplied with the matrix, the result is always an integer multiple of that vector. This integer value is the corresponding Eigenvalue of the eigenvector. This relationship can be described by the equation:

$M \times u = \lambda u$, where u is an eigenvector of the matrix M is the

matrix and is the corresponding Eigenvalue. Eigenvectors possess following properties:

- They can be determined only for square matrices.
- There are n eigenvectors (and corresponding Eigenvalues) in an $n \times n$ matrix.
- All eigenvectors are perpendicular, i.e. at right angle with each other.

The traditional motivation for selecting the Eigenvectors with the largest Eigenvalues is that the Eigenvalues represent the amount of variance along a particular Eigenvector. By selecting the Eigenvectors with the largest Eigenvalues, one selects the dimensions along which the gallery images vary the most. Since the Eigenvectors are ordered high to low by the amount of variance found between images along each Eigenvector, the last Eigenvectors find the smallest amounts of variance. Often the assumption is made that noise is associated with the lower valued Eigen values where smaller amounts of variation are found among the images [19].

C. Principle Component Analysis

Features of the face images are extracted using PCA in this purposed methodology. PCA is dimensionality reduction method and retain the majority of the variations present in the data set. It capture the variations the dataset and use this information to encode the face images. It computes the feature vectors for different face points and forms a column matrix of these vectors. PCA algorithm steps are shown in Fig. 2.

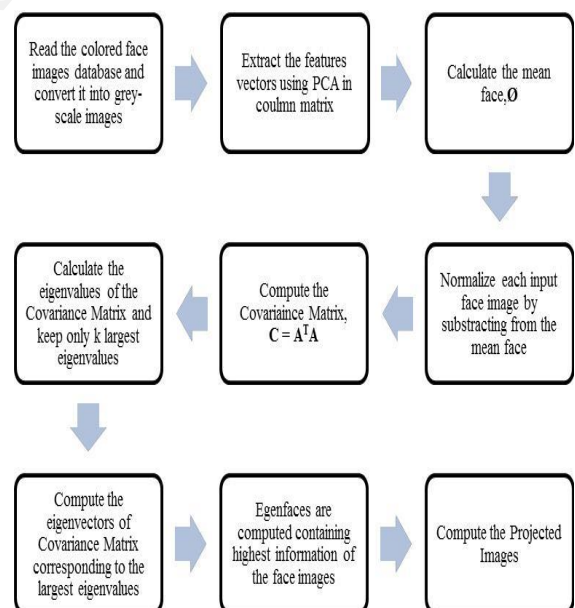


Fig. 2 Features Extraction using PCA by computing the Eigenface Images

PCA projects the data along the directions where variations in the data are maximum. The algorithm is follows as:

- Assume the m sample images contained in the database as $A_1, A_2, A_3, \dots, A_m$.
- Calculate the average image, O , as: $O = \sum A_i / M$, where $1 < L < M$, each image will be a column vector the same size.
- The covariance matrix is computed as by $C = A^T A$ where $A = [O_1 O_2 O_3 \dots O_m]$.

- Calculate the eigenvalues of the covariance matrix C and keep only k largest eigenvalues for dimensionality reduction as $\lambda_k = \sum_{n=1}^m (U_k^T O_n)$.
- Eigenfaces are the eigenvectors U_k of the covariance matrix C corresponding to the largest eigenvalues.
- All the centered images are projected into face space on eigenface basis to compute the projections of the face images as feature vectors as: $w = U^T O = U^T (A_i - \bar{O})$, where $1 < i < m$.

PCA method computes the maximum variations in data with converting it from high dimensional image space to low dimensional image space. These extracted projections of face images are further processed to Support vector machine for training and testing purposes.

D. Support Vector Machines

Support vector machines are learning machines that classify data by shaping a set of support vectors. SVMs provide a generic mechanism to robust the surface of the hyper plane to the data through. Another benefit of SVMs is the low expected probability of generalization errors. Moreover, once the data is classified into two classes, an appropriate optimizing algorithm can be used if needed for feature identification, depending on the application. SVM creates a hyper-plane between two sets of data for classification; in our work, we separate the data into two classes: face belongs to the train database and face doesn't belong to the train database. Input data X that fall one region of the hyper-plane, $(X^T \cdot W - b) > 0$, are labeled as +1 and those that fall on the other area, $(X^T \cdot W - b) < 0$, are labeled as -1.

We seek the linear classifier that separates the data with the lowest generalization error. Intuitively, this classifier is a hyper plane that maximizes the margin error, which is the sum of the distances between the hyper plane and positive and negative examples closest to this hyper plane.

We consider the example in (a) where there are many possible linear classifiers that can separate the data, but there is only one that maximizes the margin shown in (b). This classifier is termed the optimal separating hyper-plane (OSH).

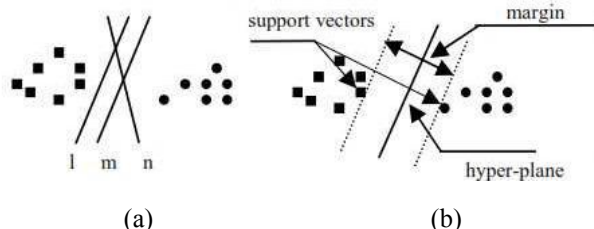


Figure 4. (a) Arbitrary hyper-planes: l, m, n; (b) Optimal hyper-plane

IV. Experimental results

We applied each feature extraction method with SVM on the ORL face database. We extracted PCA feature vectors with an application program coded using Matlab 7.0. Tests were done on a PC with Intel Pentium D 2.8-GHZ CPU and 1024-MB RAM.

In this study, standard ORL images (10 poses for each of 40 people) were converted into JPEG image format without changing their size. For both feature extraction methods a total of six training sets were composed that include varying pose counts (from 1 to 6) for each person and remaining poses are chosen as the test set. Our training sets include 40, 80, 120, 160, 200 and 240 images according to chosen pose count. For each person, poses with the same indices are chosen for the corresponding set.

For PCA based Eigenfaces approach, size of each feature vector is determined by the size of eigenface space. As the training set grows, this size reaches up to 240 (six pose for each person). For a simpler and more feasible classification process we utilize only first 40 elements of each feature vector. (In creation of eigenface space, eigenvectors were rearranged by sorting their corresponding eigenvalues.) Thus we use these 40 features for SVM.

V. CONCLUSION

In this paper, a new Face recognition method is presented. The new method was considered as a combination of PCA, and SVM. We used these algorithms to construct efficient face recognition method with a high recognition rate. Proposed method consists of Three parts: i) image preprocessing that includes histogram equalization, normalization and mean centering, ii) dimension reduction using PCA that main features that are important for representing face images are extracted iii) Support vector machine that classify input face images into one of available classes. Simulation results using YALE face datasets demonstrated the ability of the proposed method for optimal feature extraction and efficient face classification. In our simulations, we chose 10 persons and considered 40 training image and 20 test image for each person (totally 400 training and 200 test face images). Experimental results show a high recognition rate equal to 93% (in average one misclassification for each 200 face images) which demonstrated an improvement in comparison with previous methods. The new face recognition algorithm can be used in many applications such as security methods.

REFERENCES

- [1] A. Jain, R. Bolle, S. Pankanti Eds, "BIOMETRIC – Personal Identification in Networked Society", Kluwer Academic Publishers, Boston/ Dordrecht/ London, 1999.
- [2] J. R. Solar, P. Navarreto, " Eigen space-based face recognition: a comparative study of different approaches, IEEE Tran. , Systems man And Cybernetics- part c: Applications, Vol. 35, No. 3, 2005.
- [3] D.L. Swets and J.J. Weng , "Using Discriminant Eigen features for image retrieval", IEEE Trans. Pattern Anal. Machine Intel, vol. 18, PP. 831-836, Aug. 1996.
- [4] P.N. Belhumeur, J.P. Hespanha, and D. J. Kriegman, "Eigen faces vs. Fisher faces: Recognition using class specific linear projection", IEEE Trans. Pattern Anal. Machine Intel., vol. 19, PP. 711-720, may 1997.
- [5] M. Turk, A. Pentland, "Eigen faces for face recognition", Journal cognitive neuroscience, Vol. 3, No.1, 1991.
- [6] W. Zhao, R. Chellappa, A. Krishnaswamy, "Discriminant analysis of principal component for face recognition", .IEEE Trans. Pattern Anal. Machine Intel., Vol 8, 1997.
- [7] O.Deniz, M. Castrill_on, M. Hern_andez, "Face recognition using independent component analysis and support vector machines" , Pattern Recognition letters, Vol. 24, PP. 2153-2157, 2003.
- [8] B. Moghaddam, "Principal manifolds and probabilistic subspaces for visual recognition", IEEE Trans. pattern Anal. Machine Intel., Vol. 24, No. 6, PP. 780-788, 2002.
- [9] H. Othman, T. Aboulnasr, " A separable low complexity 2D HMM with application to face recognition" IEEE Trans. Pattern. Anal. Machie Inell., Vol. 25, No. 10, PP. 1229-1238, 2003.
- [10] M. Er, S. Wu, J. Lu, L.H.ToH, "face recognition with radial basis function(RBF) neural networks", IEEE Trans. Neural Networks, Vol. 13, No. 3, pp. 697-710.
- [11] K. Lee, Y. Chung, H. Byun, "SVM based face verification with feature set of small size", electronic letters, Vol. 38, No. 15, PP. 787-789, 2002.
- [12] J.J. Weng, "using discriminant eigenfeatures for image retrieval", IEEE Trans. Pattern Anal. Machine Intell. , Vol. 18, No. 8, pp. 831-836, 1996.
- [13] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd Edition, Academic Press, New York, 1990.
- [14] S. Pang, S. Ozawa, N. Kasabov, " Incremental linear discriminant analysis for classification of data streams", IEEE Trans. on Systems, Man and Cybernetics, vol. 35, no. 5, pp. 905-914, 2005.
- [15] M.J.Er, W.Chen, S.Wu, "High speed face recognition based on discrete cosine transform and RBF neural network", IEEE Trans. On Neural Network, Vol. 16, No. 3, PP. 679,691, 2005.
- [16] D. Ramaeubramanian, y. Venkatesh, "Encoding and recognition of Faces based on human visual model and DCT", Pattern recognition, Vol. 34, PP. 2447-2458, 2001.
- [17] X. Y. Jing, D. Zhang, "A face and palm print recognition approach based on discriminant DCT feature extraction", IEEE trans. on Sys. Man & Cyb., Vol. 34, No. 6, PP. 2405-2415, 2004."
- [18] Z. Pan, A. G. Rust, H. Bolouri, " Image redundancy reduction for neural network classification using discrete cosine transform", Proc. Int. Conf. on Neural Network, Vol. 3, PP. 149,154, Italy, 2000.
- [19] S. Zhao, R. R. Grigat, "Multi block fusion scheme for face recognition", Int. Con. on Pattern recognition (ICPR), Vol. 1, pp. 309-312, 2004.